
Rapport de Sciences des Données 4

TE64MI

Thème : « *Les Universités* »

Membres : **Elise Commandré,**
Alvin Ingabire,
Michèle Medom,
Berline Niyonkuru,
Grace Palenfo.

Enseignants : **Sandra Bringay,**
Namrata Patel,
Théodore Michel Picque.

1. Présentation du thème

Le thème que nous avons choisi d'aborder pour notre projet est celui des universités. Au cours de la phase de recherche d'idées pour le site, qui s'est avérée être une tâche exigeante, nous avons abouti à une conclusion unanime : *la complexité de la recherche universitaire*. Cette difficulté est devenue évidente, conduisant à la création du site « **UniDiscover** ». Il s'adresse aux récents diplômés du lycée, aux personnes cherchant à se réorienter, ainsi qu'à toute personne souhaitant obtenir des informations sur les universités.

À la différence d'autres sites d'orientation post-bac, tels que l'Onisep, UniDiscover prend en considération des aspects au-delà de la recherche de formations et de la réorientation. Notre défi principal est de permettre à toute personne cherchant des informations sur les universités de s'informer non seulement sur les programmes d'études, mais également sur le cadre de vie des villes abritant ces établissements.

Nous nous sommes lancés dans la recherche de jeux de données informatifs et fiables pour garantir l'authenticité et la fiabilité du site. L'objectif de cette recherche était de fournir à notre futur site et à ses utilisateurs des informations sur :

- Le coût de la vie dans les villes répertoriées, notamment le logement et les courses alimentaires, entre autres.

- Le taux de criminalité.

- Le classement des universités et les formations qu'elles dispensent

Il s'est avéré que la collecte d'informations sur la criminalité et le coût de la vie dans toutes les villes, au niveau mondial, est quasiment impossible. Cela est dû au manque de données complètes dans de nombreux pays. Nous avons donc décidé de nous concentrer sur les États-Unis, qui disposent du plus grand nombre d'informations conformes à nos attentes. Ainsi, UniDiscover se concentrera sur les universités américaines.

2. Sciences de données

Dans cette partie du semestre, nous avons été amenés à coder plusieurs pages.

Pour la page ***predict***, nous avons utilisé la régression linéaire pour pouvoir prédire le rang des 50 premières universités pour l'année à venir. Ce processus d'entraînement sur des données de notre base de données s'est fait par plusieurs étapes :

- l'utilisation de deux variables X et y pour stocker les variables. X renfermait les différents paramètres sur lesquels le modèle de prédiction s'entraîne et y constituait la variable à prédire.

-Au vue des disparités observées sur les 50 premières valeurs de nos données par rapport au reste de nos données, nous avons décidé de ne garder que celle-ci pour maximiser au mieux, les résultats de ce classement.

-Nous avons utilisé la `LinearRegression()` de Scitkit-Learn pour entraîner le modèle et avons utilisé le Coefficient de Détermination (R^2) pour estimer la viabilité du modèle. Nous avons obtenu un R^2 de 58% ce qui n'est pas très élevé mais assez pour rendre justifiable nos résultats.

Pour plus de détails sur le code et les méthodes utilisées, se référer au notebook "Rapport_sciences_de_données4.ipynb".

3. Description des fonctionnalités

-Accueil.php

- Utilisateur non connecté :
 - Se connecter / Créer un compte
- Utilisateur connecté :
 - Suggestions personnelles s'il a répondu au questionnaire après l'inscription
 - Navigations sur les différentes pages à l'aide des onglets sur le bandeau

-Compare.php

-Permet de faire le choix de comparer soit deux universités entre elles soit deux villes entre elles.

- redirige vers **comparaison_unis.php**: permet de visualiser le rang des universités et les scores_teaching et scores_international_outlook et aussi de comparer les différents scores et l'évolution du rang de deux universités différentes
- redirige vers **ville.php**: qui permet de comparer la qualité de vie et la criminalité deux villes

-Localisation.php

- Permet de visualiser géographiquement l'emplacement des universités, et en cliquant sur le point géographique d'une université, de renvoyer sur la page de l'université correspondante.

-Predict.php

-Affiche le classement prédit des universités pour l'année d'après

-Contact.php

-Présentation des créateurs sites et possibilité de les contacter pour retours et renseignements

-Search.php

-Permet de rechercher des universités, et de renvoyer sur la page de l'université correspondante avec une présentation de celle-ci

-La page de l'université, universite.php affiche une description de l'université, et un graphique permettant de visualiser l'évolution du classement de l'université entre 2016 et 2023. Le graphique a été réalisé avec chart.js.

-Favoris.php

-Affiche la liste des université passées en favoris depuis la page université.php

-monCompte.php

-Affiche et récapitule le profil de l'utilisateur connecté
-Possibilité de se déconnecter

Description du But et des Étapes de Développement de la Page

search_university.html:

La conception de la page search_university.html a pour objectif de fournir aux utilisateurs un moyen simple de rechercher des universités américaines en fonction de leur nom. Dans cette optique, nous avons élaboré une interface intuitive, permettant aux utilisateurs de saisir le nom d'une université dans un champ de recherche et d'obtenir des résultats en temps réel.

Nous avons commencé par établir une connexion à la base de données contenant les informations sur les universités américaines, puis nous avons procédé à la

récupération du terme de recherche saisi par l'utilisateur dans le champ de recherche de la page.

Parallèlement, nous avons intégré un code en JavaScript pour offrir des suggestions à mesure que l'utilisateur commence à saisir dans le champ de recherche. Cette fonctionnalité améliore l'expérience utilisateur en fournissant des options de recherche dynamiques.

Une fois les résultats obtenus, nous les avons affichés sur la page HTML à l'aide de JavaScript. Les résultats de la recherche sont présentés sous forme de liste déroulante, permettant à l'utilisateur de sélectionner facilement une université parmi les suggestions proposées et en cliquant sur le nom de l'université de se rendre sur la page correspondante.

Description du but de la page ville.php

La page ville.php permet aux utilisateurs de comparer la qualité de vie et la criminalité entre deux villes américaines. Cette fonctionnalité repose sur notre base de données contenant des informations détaillées sur le coût de la vie et la criminalité dans diverses villes aux USA.

Après avoir choisi les villes à comparer, on a trois graphiques interactifs (générés à l'aide de chart.js et un tableau qui permettent de visualiser la comparaison entre les deux villes :

- le premier graphique compare le coût total de la vie des deux villes calculés à partir de la formule suivante:

$$2 * \text{Meal} + 3 * \text{McMeal} + 30 * \text{Capuccino} + 10 * \text{CokePepsi} + 4 * \text{Milk} + 30 * \text{FreshBread} + 5 * \text{Rice} + 15 * \text{Eggs} + 10 * \text{LocalCheese} + 5 * \text{ChickenFillets} + 3 * \text{BeefRound} + 5 * \text{Apples} + 5 * \text{Banana} + 5 * \text{Oranges} + 5 * \text{Tomato} + 5 * \text{Potato} + 5 * \text{Onion} + 2 * \text{Lettuce} + 10 * \text{Water} + 2 * \text{DomesticBeer} + 3 * \text{Cigarettes20Pack} + \text{MonthlyPass} + 30 * \text{Gasoline} + \text{Internet} + \text{FitnessClubMonthly} + 4 * \text{Cinema} + \text{OneJeansLevis} + \text{SummerDress} + \text{PairOfNike} + \text{PairOfMenShoes} + \text{AppartmentOneBedroomCity} + \text{BasicFor85m2}$$

- le second compare la criminalité des villes avec un indicateur calculé à l'aide de la formule suivante :

$\text{sum}(\text{crime.victims_killed} + \text{crime.victims_injured} + \text{crime.suspects_killed} + \text{crime.suspects_injured} + \text{crime.suspects_arrested})$

- le troisième compare l'indice des coûts de vie en 5 catégories (le prix de l'alimentation, des services, du logement, du shopping et les salaires mensuels net) avec les formules suivantes :

☐ **alimentation:** $\text{Meal} + \text{MealFor2_people} + \text{McMeal} + \text{Capuccino} + \text{CokePepsi} + \text{Milk} + \text{FreshBread} + \text{Rice} + \text{Eggs} + \text{LocalCheese} + \text{ChickenFillets} + \text{BeefRound} + \text{Apples} + \text{Banana} + \text{Oranges} + \text{Tomato} + \text{Potato} + \text{Onion} + \text{Lettuce} + \text{Water} + \text{DomesticBeer}$

☐ **services:** $\text{TaxiStart} + \text{TaxiUnKm} + \text{Internet} + \text{FitnessClubMonthly} + \text{OneWayTicket} + \text{MonthlyPass} + \text{Cinema} + \text{OneMminPrepaid}$

☐ **logement:** $\text{ApartmentOneBedroomCity}$

☐ **salaires:** $\text{AverageMonthlyNetSalary}$

☐ **shopping:** $\text{PairOfNike} + \text{PairOfMenShoes} + \text{SummerDres}$

- Le dernier tableau récupère le coût de quelques produits et services, le logement et le salaire puis montre la différence entre les deux coûts en %.

On a aussi des commentaires pour les deux premiers graphiques et le dernier tableau pour aider l'utilisateur dans l'interprétation des graphiques et des données du tableau.

On a ainsi une page fonctionnelle et intuitive qui permet aux utilisateurs de comparer les différents aspects du mode de vie entre les deux villes.

4. Conclusion

4.1. Difficultés rencontrés

En ce qui concerne la page prédire (prediction.html), le plus gros challenge a été au niveau de l'entraînement sur un modèle de régression linéaire. Les variables dans la table classement n'avaient pas les mêmes tailles plus nous redescendions vers les années récentes. Le modèle à une accuracy (58%) plutôt moyenne due au fait que nous ayons dû transformer les classements 'rank_order' en score et ensuite après la prédiction les retransformer en rang.

4.2. Bibliothèques et technologies utilisées

-Scitkit-learn

-Flask

-Python

-Mysql

-Json

-JavaScript (chart.js, mapbox.js)

-YouTube

-ChatGPT

-API

5. Liens:

Lien github: https://github.com/BGHniyonkuru/Projet_Universite

Lien vers le site: www.unidiscov.go.yj.fr

Lien vers la vidéo: https://www.youtube.com/watch?v=nNUtPjwf_Qo